# The Battle of Neighborhoods:

# Starting a Cloud Kitchen Business

## A. INTRODUCTION

### A.1. Background

Jakarta is the capital of Indonesia. It is the center of economy, culture and politics in Indonesia with a population of 10,770,487 in the city as of 2020. It is also the world's second-most populous urban area after Tokyo. Jakarta offers great business opportunities, as well as a higher standard of living, that have attracted migrants from across the Indonesian archipelago or even immigrants across countries. Financial institutions and corporate headquarters of numerous Indonesian companies and multinational corporations are located in the city, making it full of high rise buildings. Jakarta's GRP was estimated at US$483.4 billion in 2017.

Food & beverage (F&B) business sector plays an important role to gross domestic product (GDP) of Indonesia. It has steadily increased since in the past five years, reaching 6,77% contribution by 2019. As in 2020, this sector has been hit hard by the COVID-19 pandemic as people opt to stay in rather than going out for meals. Long before the pandemic, Jakarta is challenged by gridlocked traffic, congestion, forcing people to spend most of their time on the road or commuting. Restaurants and cafes have been dealing the reduction of dine-ins by providing online delivery service, in addition of the rise of ride hailing start-ups which also provide similar services. This condition creates the opportunity for cloud kitchen or virtual kitchen to emerge as a potential business model.

Cloud kitchen is a professional food preparation and cooking facility set up for the preparation of delivery-only meals. To simplify, cloud kitchen is a restaurant without front-of-house operation. This concept allows restaurants to operate without having a physical presence at a central hip location, solving the problem of high property cost and initial set up cost. Consumers are also have the benefits to explore wide range of choices of food and beverage with delivery services.

### A.2 Problem

It is not hard for people in Jakarta to find varieties of food and beverage around the city, and this might be a problem if we want to determine the best place to open a cloud kitchen. The objective of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the ‘best’ neighbourhood or district in Jakarta to open a cloud kitchen facility. Ideally, we want to find districts which don’t have restaurants or cafes as their common venues.

### A.3. Target Audiences

This project is targeted for entrepreneurs who are interested about the concept of cloud kitchen and planning to open a cloud kitchen facility. This project is also targeted for business owners and stakeholders who want to expand their business and interested about how data science is applied to solve this kind of problem.

## B. DATA ACQUISITION AND CLEANING

### B.1 Data Description

The followings are data sources that is used for this project:

* List of administrative cities, districts and sub-districts in Jakarta (excluding Kepulauan Seribu). This information can be found in Wikipedia
* Coordinates of each districts in Jakarta. This data is prepared by the author.
* The popular or most common venues of a given district in Jakarta. This information can be found inside Foursquare Location data, using Foursquare API to access it

To simplify, we will load the list of districts in Jakarta. Then, we will add the coordinates for each of the districts. Finally, using the coordinates of the district, we will use Foursquare credentials to acess the venues around them with their details. The frequency of the venue category in each neighborhood will be the features of the clustering model.

### B.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There are several problems with the datasets. The lists of districts are scattered between different pages of Wikipedia based on the administrative cities they belong to. We have to scrape them individually using BeautifulSoup libraries and transform it into dataframe. For example, below are the steps of scraping table of disctrits in Jakarta.

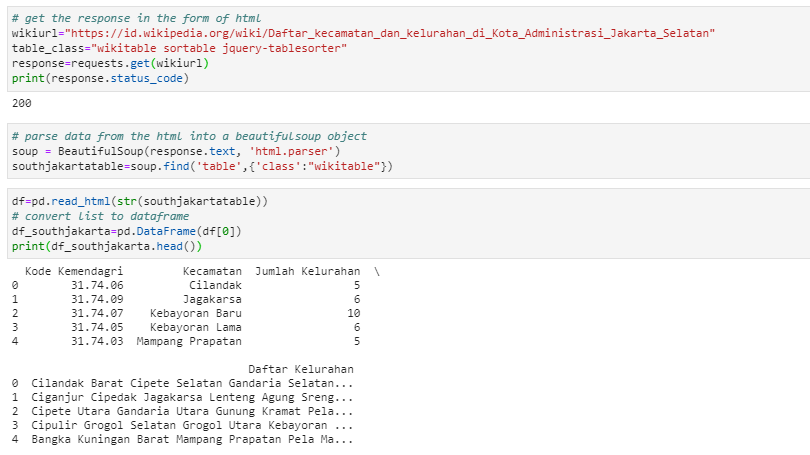


Figure 1 Scraping Districts in South Jakarta from Wikipedia

We will repeat the steps for other administrative cities and concatenate the dataframes into one table.



Figure 2 First Five Rows of Districts in Jakarta Dataframe

We don’t need the “Kode Kemendagri” column. Therefore we should drop those column.

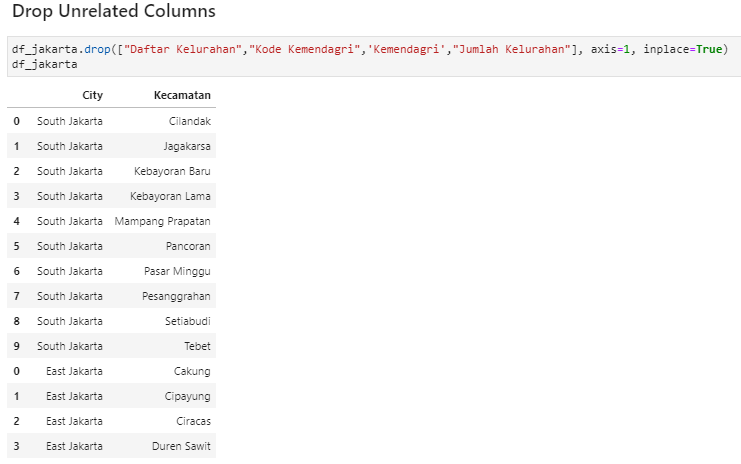


Figure 3 Dropping Unrelated Columns

The author has already prepared a file containing districts coordinates. We should read the file and transform it into dataframe.

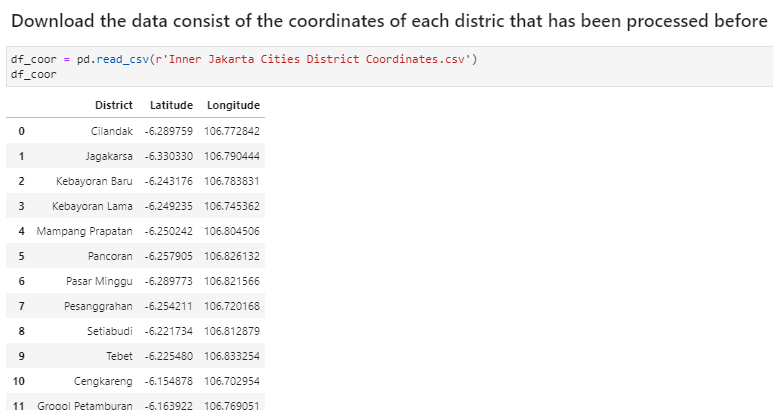


Figure 4 Coordinates of Districts in Jakarta

Lastly, we concatenate the districts dataframe with the coordinates dataframe.

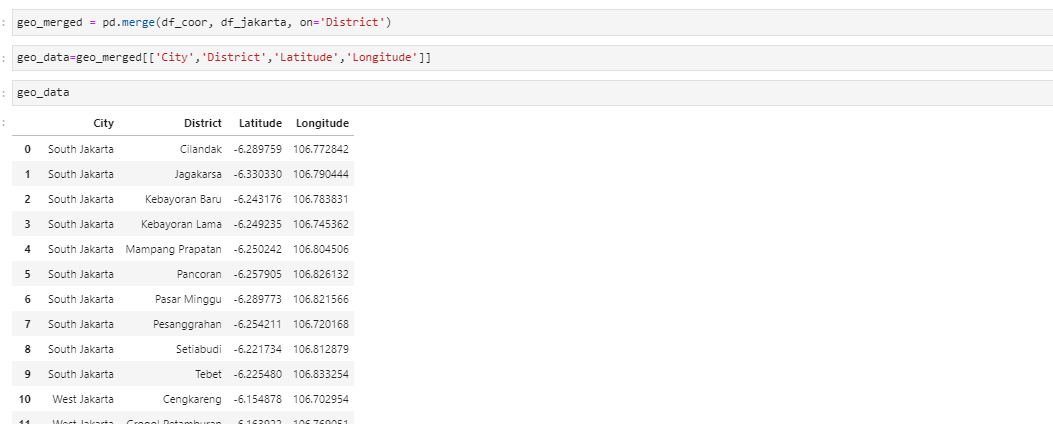


Figure 5 Districts Coordinates in Jakarta

The data is now ready to be processed.

## C. METHODOLOGY

### C.1. Analytic Approach

We approach the problem using the clustering technique, namely ***k.Means*.** This approach enables the audience to see how similar districts about their demographics. We can then examine each cluster and determine the discriminating venue categories that distinguish each cluster. After that, we can decide the best cluster where we will open the cloud kitchen facility.

### C.2. Exploratory Data Analysis for Districts Data

We have built a district dataframe that contains **42 districts, 5 administrative cities** where the districts belong to, and **their coordinates.** We will do an exploratory data analysis.



Figure 6 The first 5 districts in Jakarta

We will use the **folium** library to visualize geographic details of Jakarta and its 42 districts using the coordinates.

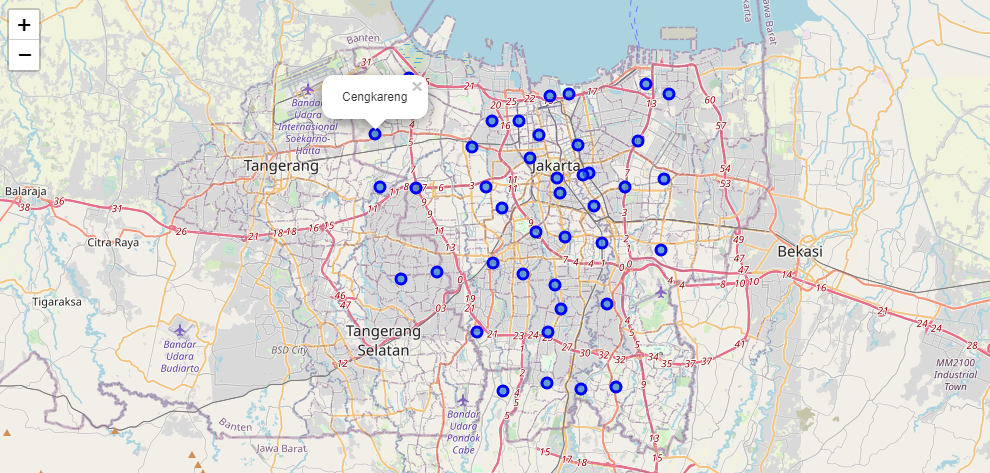


Figure 7 The district distribution in Jakarta

Given the coordinates information of each district, we can use the Foursquare API to access top 100 venues in each districts of Jakarta within a radius of 500 meters.

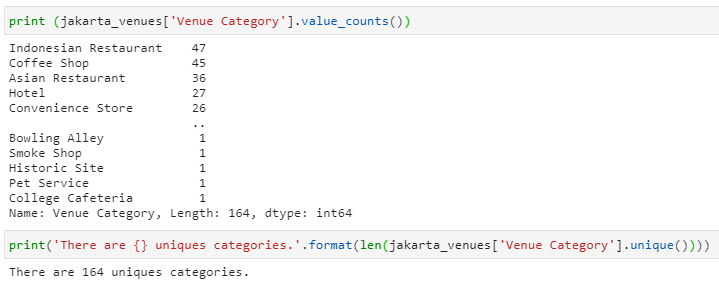


Figure 8 Identifying unique venue categories

We can tell from the picture above that there are 164 unique venue categories. Next, we will find out what are the top 10 venue categories occurring in districts in Jakarta using barchart.

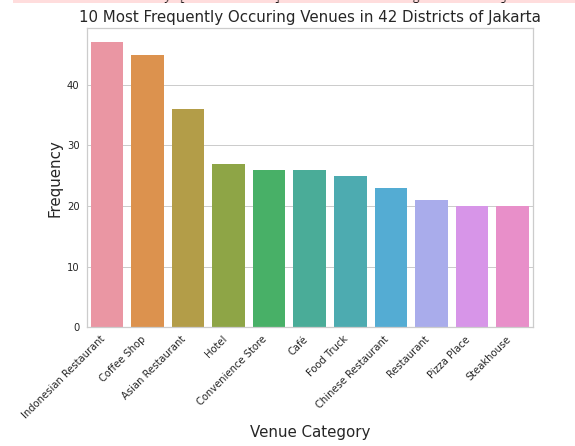


Figure 9 Top 10 venue categories in Jakarta

As we have assumed before, it is not hard for people in Jakarta to find food and beverage places, especially Indonesian restaurants, Asian restaurants, and coffee shops. This will pose a problem about where we have to establish our new cloud kitchen facility. We will have to find which district that are not too populated with restaurants and coffee shops using the clustering model.

### C.3 Clustering the Districts

We will run the **k-Means** algorithm to build a clustering model with a different number of cluster (k). The features will be the means of the frequency occurrence of each venue category. First, we have to transform the venue categories data into numerical data so that the ***k-Means*** algorithm can process them. We transform the data using **one hot encoding**.



Figure 10 One hot encoding

Notice that the number of the column is 165, similar to the number of the venue categories plus the district column. Next, we will use pandas groupby on district column and calculate the mean of the frequency of occurrence of each venue category.



Figure 11 Grouping venue category based on occurences in each districts

After that, we will determine the most common values in each districts sorted by the frequency of the occurrence.

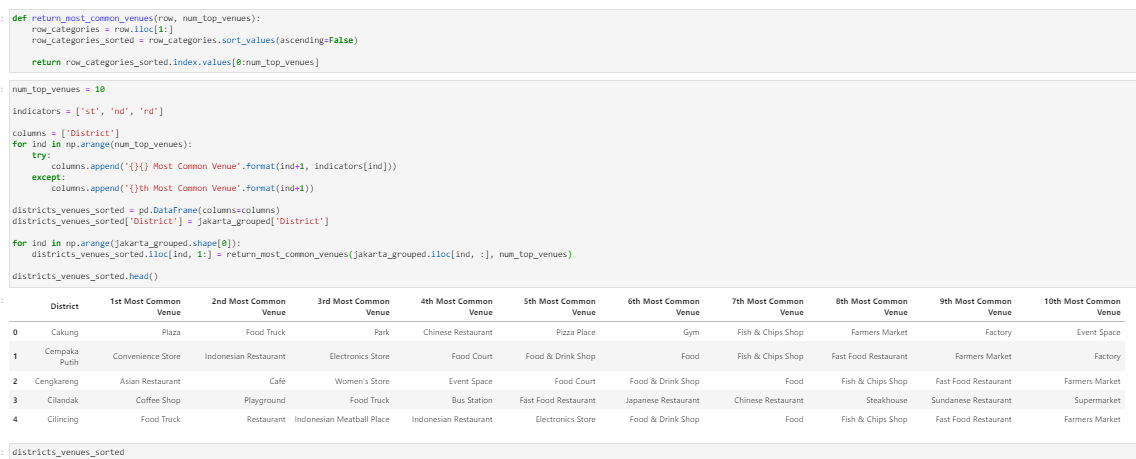


Figure 12 Most common venues in each districts

Finally, we can use the data provided by the above dataframe to cluster the districts. To determine the best number of **k**, we will use **KElbowVisualizer** algorithm.

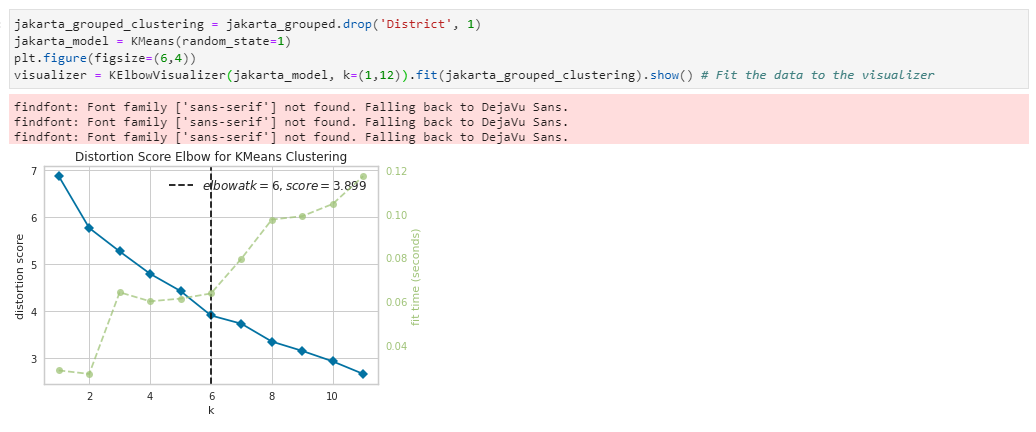


Figure 13 KElbowVisualizer algorithm

We can tell from the picture above that the best number of **k** is 6. Hence, we will have 6 clusters at the end of this project. Let’s run the clustering model with 6 as the number of **k** and visualize it with map.



Figure 14 Code clustering and visualizing districts

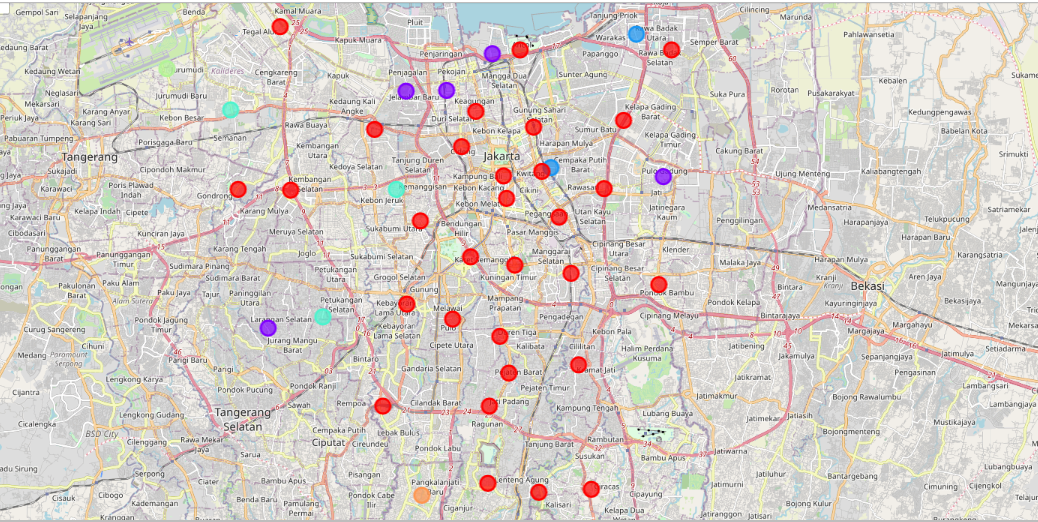


Figure 15 Clusters of districts in Jakarta

We can then examine venues listed inside each cluster and define the discriminating venue categories that distinguish them.



Figure 16 Most common venues of cluster 1



Figure 17 Most common venues of cluster 2



Figure 18 Most common venues of cluster 3

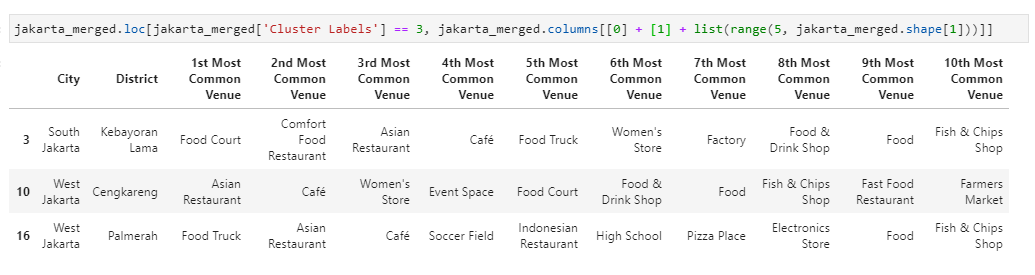


Figure 19 Most common venues of cluster 4



Figure 20 Most common venues of cluster 5



Figure 21 Most common venues of cluster 6

## D. RESULTS AND DISCUSSIONS

We have used the **k-means** algorithm to cluster the districts in Jakarta, resulting in 6 clusters with the information of most common venues of each district. The clusters are:

* Cluster 1: Hangout Venues. Many of the districts in this cluster have similarities for their first, second, and third most common venues such as coffee shop, café, restaurant, and food truck. The cluster also consists of districts that are considered as hip place among local people such as Cilandak, Kebayoran Baru, Setiabudi and Tebet.
* Cluster 2: Gathering Venues. The districts in this cluster have similarities for their first most common venues such as plaza and playground. Those are kind of venues where people usually gather together to meet up.
* Cluster 3: Convenience Store Venues. The districts in this cluster have similarities for their most common venues, which is convenience store. It is possible that there is high density of housing complex in the districts.
* Cluster 4, 5, and 6 have similarities in their common venues, such as restaurants, food trucks, and food court. They are divided because of the varying frequencies of venue categories in each districts.
* Cluster 2 and 3 are recommended districts for further inspection. We can use other factors, such as population density or average income, to further analyze and deep dive into each distric based on target market of the cloud kitchen.

The clustering is completely based on the most common venues obtained from Foursquare data. The real condition could be different as there might be place that have not yet included in Foursquare. We have also ignored other factors like demographics of citizens, average income, property rental price, and so on, since we don’t have such data and it would be difficult to process it for a preliminary study like this project. Hence, our analysis only helps target audiences to get an overview of districts clusters based on venue categories in the districts. Furthermore, this results also could potentially vary if we use some other clustering techniques.

## E. CONCLUSION

Determining location to open a new business could be challenging due to many factors affecting it. Cloud kitchen is a new concept and we have to be careful to assess the factors affecting this business. However, this project gives meaningful insights about how districts in Jakarta is clustered based on real data. This will help the target audiences to further analyze how to determine the best location to open cloud kitchen facility or other problems that requires geospatial data and clustering.

## F. REFERENCES

* <https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Selatan>
* <https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Barat>
* <https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Utara>
* <https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Timur>
* <https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Pusat>
* <https://en.wikipedia.org/wiki/Ghost_kitchen>
* <https://medium.com/locale-ai/operational-efficiency-for-cloud-kitchens-using-location-data-aab5f0a796b7>
* <https://www.thejakartapost.com/news/2020/03/27/food-beverage-industry-hit-hardest-by-covid-19-report.html>
* <https://www.thejakartapost.com/news/2019/07/02/fb-contribution-to-gdp-steadily-increases.html>